Executive Summary

LEAD SCORING CASE STUDY

Pritam Panchal & Padmanabhan M. | DSC-58 | Date: 15/01/2024

Introduction:

This case study is for an education company named X Education that sells online courses to industry professionals. X Education has appointed us to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company requires us to build a model wherein you need to assign a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance. The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

Approach and Methodologies:

1. Import the datasets & libraries
   Imported the dataset and all the required libraries for the analysis and model building.

2. Check for general structure of the data.
   Checked the basic metadata/structure of the data to have a better understanding of the data.

3. Handling missing values & Outliers (Data Cleaning).
   Most of the columns with high missing values above 40% were dropped and others were replaced with the mode of the columns and at the end we had the data with no missing values. There were no major outliers in the data as well.

4. Analyze using Univariate and Bivariate Analysis.
   Through the analysis we found the variables and its correlation with the target variable and also removed unnecessary variables.

5. Creating dummies for the categorical variables.
6. Splitting data into train & test set and further performing Feature Scaling & Selection.
7. Building a Logistic Regression Model until we get a good balance of parameters.
8. Model Evaluation
   According to the optimal cutoff of 0.35, the accuracy, sensitivity & specificity came out to be around 80%.

9. Making predictions on test set and evaluating again & assigning a Lead Score.
   After making the predictions on the test dataset, we achieved 80% accuracy, sensitivity & specificity

Final Model Results:

Train set
- Accuracy : 80.94%
- Sensitivity : 81.55%
- Specificity : 80.56%

Test set
- Accuracy : 80.86%
- Sensitivity : 82.32%
- Specificity : 80.04%

Hence, we have achieved a lead conversion rate of 80%.

Insights & Recommendations:

After analyzing the datasets, we found the variables in the model which contribute most towards the probability of a lead getting converted:

1) Page Views Per Visit
2) TotalVisits
3) Total Time Spent on Website
4) Page Views Per Visit

- The team should make phone calls if the lead has visited the website multiple times and also spend more time on it. To increase this the website should be built more intuitive and engaging.
- Leads from source Welingak Website and having current occupation_Working Professional should be kept a preference.
- Company should make a cutoff of the lead score and contact only the leads above the cut off score.
- To reduce calls, company can send automated emails and SMS system to keep the interaction live with the hot leads.